We will also measure several different factors over an extended period to determine the effect of our system on consensus among experts in the field. First, we will keep track of the rate of changes made to the knowledge base, in terms of both the growth (addition of new material) and modifications (changes to existing material). Our premise is that a decrease in changes (perhaps even to zero) indicates that the experts using the system have come to an agreement on the basic decision criteria to be used and the appropriate answer for each case.

We will measure the completeness of the knowledge base by the number of counterexamples (proposed either by the experts, or perhaps ultimately by the system itself) that force addition of new rules or changes to existing rules.

While a decrease in changes to the knowledge base and number of counterexamples may suggest a consensus has been reached, it is important to verify that the agreed-upon set of decision criteria is in fact correct. For this reason, we will also monitor the correctness of the knowledge base by evaluating the quality of MYCIN's conclusions. This will be done by asking other experts to rate the appropriateness of MYCIN's conclusions and recommendations.

This will also help us to measure the variability between experts. In infectious diseases, as in any other growing discipline, there is still some disagreement among experts as to what the 'best' recommendations should be. We will measure this variability by proposing several cases to a panel of experts and asking for their opinions about MYCIN's and each others' recommendations. This measure will be important in determining the level of consensus before and after interaction with our programs. A decrease in this inter-expert variability will provide an indication that interacting with our system compels the expert to recognize explicitly the criteria that should be employed in reaching a decision, and hence provides an effective forum for discovering variations in those criteria among experts.

5   Significance of the Research

By assembling the program's knowledge base of rules, we will arrive at a compilation and systematization of the current knowledge of infectious disease diagnosis and therapy. While any one expert may be able to supply only a part of that entire collection, by calling on the services of many experts it should be possible to construct what may become a unique reference source for currently accepted practice.

A system such as MYCIN can provide a source of consistent, up-to-date consultative advice, available at all hours to any

physician with a computer terminal and a telephone. It can be systematically modified to reflect regional differences in clinical practice, and quickly updated to take advantage of progress in medical research. We believe that in the long run it can favorably affect the prescribing habits of physicians, resulting in better medical care.

In addition, the system may have a significant educational impact. It is prepared to offer a detailed explanation for every step in its diagnostic process, and can also answer more general questions about its knowledge of the field. These explanation and question answering capabilities not only assure the clinician that the program reaches its conclusions by a reasoning process similar to his own, but can provide a strong instructive influence for the student.

Finally, where most attempts at quality assurance are retrospective and involve mechanisms like chart review, MYCIN offers the possibility of prospective assistance. This is not only effective in maintaining quality, but by offering assistance before treatment is initiated, can have a more immediate impact on health care practice. Prospective intervention is also likely to meet with greater physician acceptance, since it offers him an opportunity to obtain advice before acting, encouraging him to avoid making mistakes rather than pointing them out after the fact.

MYCIN may also be useful in situations where chart review remains the preferred technique for quality assurance. A common problem with the standard approach is that it requires either subjective judgments and a significant time investment by the very specialists whose expertise is in short supply, or the use of a single set of global criteria by which to evaluate performance, promoting what has been called 'stereotyped medicine'. The existence of a program whose performance was known to be of high quality would provide an effective solution. House staff could conduct the chart review (freeing the specialist), and the system would provide a perfectly repeatable, objective standard by which to judge performance. Note that MYCIN is currently capable both of making specific conclusions on the basis of each case individually, and of offering an assessment of the range of possible causative organisms and therapeutic regimens. It thus becomes possible to evaluate performance on individual cases, rather than setting global (usually statistical) standards, and to judge the accuracy of a range of answers.

6   Facilities Available

The Stanford University Medical Experimental Computer (the

27

SUMEX-AIM system) is a dual-processor, time-shared Digital Equipment Corporation PDP-10 available via both a number of direct dial phone lines and the TYMSHARE national network of telephone lines. The system is a National Biotechnology Resource for applications of Artificial Intelligence to Medicine (AIM). MYCIN is one of the research projects accepted as part of the national AIM community, and given access to the system at no cost. Since all of MYCIN's development for the past three years has been on the SUMEX system, this represents a significant saving.
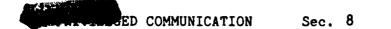
The Stanford University Medical Center and Computer Science Department and the University of Arizona Medical Center are both involved in this work as a result of the participation of the co-principal investigators and Dr. Stanton Axline and associated clinical fellows. As noted above, we have used the Stanford Center as a source of both cases on which to test the system and physicians who can evaluate its performance, and will involve the faculty and fellows of both Centers in ongoing development and evaluation programs.

7   Collaborative Arrangements

Dr. Axline has been a part of the project since its earliest days, and along with the principal investigator functioned as co-principal investigator during the initial three years of our work. He will continue to direct the University of Arizona portion of the project, acting as a primary source of infectious disease expertise, to improve the performance of the system. In addition, he will help design and carry out our evaluation program. This will offer the added benefit of giving us two different clinical groups contributing to knowledge base development, as well as a new patient population for program evaluation.

The grant to SUMEX makes explicit the importance of collaborative scientific work, and to further this the SUMEX staff have provided a number of support facilities that make joint work more feasible. One of them is a collection of message handling programs which make communication from remote sites quite easy. Other facilities make it possible for one user to run a program while another user (anywhere else in the country) 'watches over his shoulder,' perhaps offering a advice and evaluation.

We expect as a result of all these factors that continued collaboration with Dr. Axline will offer significant advantages.

8   Appendix A: Progress Report Submitted to BHSRE

8.1   Summary

Over the past three years we have designed, built and partially evaluated a computer program capable of diagnosis and therapy selection for certain varieties of infectious diseases. The program is intended to function as a consultant, and 'interviews' a doctor about his patient, requesting information on clinical findings and results of laboratory tests. It relies on a store of judgmental knowledge (obtained from experts in infectious disease) to determine the conclusions which can be drawn from the answers it receives. This judgmental knowledge is in the form of some 400 decision rules dealing with the wide range of topics that must be considered in determining the likely identity of causative organisms and selecting appropriate antimicrobials.

MYCIN is composed of the three systems described earlier (the consultation, explanation, and knowledge acquisition systems), all of which reference the knowledge base of decision rules. The program is currently capable of dealing with bacteremia and meningitis infections. It can diagnose the likely presence of more than 35 different organisms and can recommend therapy for 100 organisms, selecting drugs from a 'pharmacopoeia' of 30 antimicrobials. The system can tailor its therapy recommendations to a specific organism and infection, can adjust dosage levels and durations in response to impaired renal status, and can combine drugs to create combination therapies, giving it a wide range of clinical applicability.

8.2   Detailed Report

Our work in the past several years has been organized around five main areas of investigation. We have

a) increased the system's competence in existing areas of clinical expertise while expanding its scope

b) developed a number of user-oriented features to increase the program's attractiveness to clinicians

c) developed a range of knowledge acquisition capabilities to speed the process of expanding the system's clinical competence

d) solved a number of technical problems to insure that the program does not outgrow the computer resources available to it

e) evaluated the system's level of expertise.

8.3   Clinical Capabilities

Since the primary qualification for any clinical consultant is competence in the domain, we have devoted significant effort to expanding MYCIN's knowledge base and widening its scope of competence.
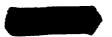
For instance, the system was directed initially at patients with positive blood cultures, the basic methodology was generalized to support a much broader approach to the problem. MYCIN has now gained the ability to deal with infections from which the causative   pathogen hasn't been   isolated (e.g., pneumonia), or which haven't even been cultured (e.g., brain abscess). With this broadening of scope, it has also become necessary to be able to evaluate the meaningfulness of isolates for cultures taken from sites other than blood. For urine and sputum isolates, for example, the system gained the ability to base its evaluation of sterility of an isolate on both the method of collection and the user's estimation of conscientiousness of collection.

An extensive review of the program's approach to drug selection has led to a major revision in the basis for therapy selection during the course of program development. The program was given the ability to consider both the infectious disease diagnosis and the significance of the organism as further determinants of therapy, in addition to organism identity. These three together have become the primary factors in drug selection, with drug toxicity and ecological factors as secondary considerations. The result is a more appropriate, more sharply focussed drug selection that also includes dose, route, and duration.

While the initial development of the knowledge base focussed on rules concerned with the diagnosis and therapy for blood infections (bacteremia), the complexity of infectious disease therapy and the frequent occurrence of multiple infections in a single patient requires a broader knowledge if the system is to be clinically useful.

In response we have extended MYCIN's knowledge base, while at the same time improving the degree of sophistication with which the system deals with bacteremia. The second major area has been the diagnosis and treatment of meningitis, and more than 100 rules were added to provide the ability to deal with it. In the processs the program was also extended beyond bacteria, as it gained the ability to consider and treat both fungi and viruses.

This area has proved to be an especially useful domain

because it has presented several new challenges. In particular, meningitis requires the ability to deal with a disease that is often diagnosed on clinical grounds alone, before any specific microbiological evidence is available (by comparison, the diagnosis of bacteremia on clinical grounds alone is far less certain, and usually requires establishment of the fact that bacterial growth has occurred in blood cultures.) For this reason, extension of the project into the meningitis area has made it necessary for MYCIN to consider a larger range of clinical factors, and has resulted in a system which has a broader picture of the whole patient.

Other contributions to the system's competence have come from expansion of the knowledge base to include information about normal bacteriological flora for a wide range of culture sites. This enables the program to distinguish between normal and pathological flora, and it can as a result decide more precisely on whether to treat.

8.4   User Oriented Features

Clinicians traditionally shun computer programs, and we believe this is in large measure due to insufficient attention paid to user oriented features. As a result, we have devoted significant effort to insuring that MYCIN is responsive to its users in a number of unique ways. The development of the explanation and question answering capabilities have been a essential for this work, and both have grown extensively in power.

The system's ability to explain the motivations for its questions, for instance, underwent a major design revision. It is now based on a more powerful approach that relies on the program's knowledge of its own control structure and ability to examine its own rules. The user can now fully explore the system's current line of reasoning, rather than just a single level, as initially implemented.

The language understanding capabilities of the question answering system have also been extensively revised. They now allow a broader range of questions to be asked and offer more precise answers. The use of this feature was also simplified so that the user no longer needs to classify his questions.

A comprehensive review of the kinds of questions asked by users of the system has led to a number of important features. MYCIN can now answer a much wider range of questions, and can, in particular, explain why it did not take a specific action, as well as why positive conclusions were reached. It is our feeling that capabilities such as these are of great importance in enabling the project's staff and clinical experts to understand

the program's rationale for its actions in instances where its recommendations do not appear to be the most appropriate and most correct. Thus, the line of reasoning of the program can be evaluated, and requirements for new or modified rules can be uncovered. These kinds of capabilities are also important in optimizing user acceptance of the system.

A substantial addition to the question-answering facility enables the system to explain the process of therapy selection. In comparison to the diagnostic process, therapy selection is complicated somewhat by the need to consider a range of different factors simultaneously, such as the total number of drugs recommended, the degree of sickness of the patient, possible interactions between drugs, toxicity and other side effects, etc. Despite this complexity, explanations of therapy selection are phrased at a conceptual level that makes them comprehensible to the physician. As before, this makes it possible for the physician to verify the validity of the system's decisions, and makes it clear to him that the system reaches its results in much the same way that he does.

The explanation consists of a step-by-step review of the reasoning which led to recommending a particular drug for a specific organism. It considers such issues as why a drug was first considered for an organism, why a drug may have been chosen as the best therapy for that organism, how the total number of drugs was reduced by considering common drug classes among the candidates, and consideration of possible contraindications based on the patient's allergies, age, and other factors. By characterizing each drug according to this scheme, the program can explain why a drug was or wasn't prescribed, as well as why one drug is to be preferred over another. This offers an important explanatory capability that will make the system more attractive and acceptable to clinicians.

Several capabilities have been added to make the program easy to use. The system is now more tolerant of erroneous or inappropriate responses, and is able to provide a reworded question, along with a list of acceptable answers. In addition, it has the ability to recognize responses which are not sufficiently precise, and can rephrase its questions accordingly.

We have recently added to the system the ability to modify drug dosage in cases of renal failure. Where, previously, the system only issued a warning to modify doses, it is now able to use either creatinine clearance or serum creatinine levels to compute the level of renal function. The program then uses drug-specific information (e.g., half-life, percent loss of the drug via renal excretion, etc.) to adjust the regimen. It can either (a) adjust dose levels downward and leave dosing interval unchanged, or (b) increase dosing interval and leave levels unchanged, or (c) allow the physician to select a dose interval, for which it chooses an appropriate dose level.

Since the problem of determining renal status and the proper adjustment of drug dose is important in the use of aminoglycoside antibiotics, cephalosporins, and other antimicrobial agents, the customization of drug dosage recommendations will be an important addition to the power of the system.

We have found, in addition, that there is a substantial amount of information that is routinely collected in every consultation, like the date and site of each of the cultures, gramstain and morphology results for each of the organisms that grew out, etc. Currently, the program exhaustively analyzes each culture and all of its organisms in turn. Some users of the program appear to be impatient with this method, and would much prefer to enter all the relevant data on all the cultures and organisms at once. This is faster and easier, since the information can be gathered in a single review of the chart, instead of having to review it several times as each culture is processed. In response to this, we have reorganized the consultation slightly, so that it is possible to enter all of this data at once, at the beginning. This offers two other advantages in addition to improving the program's acceptability to its users. First, it provides a basis for our future efforts to write rules which deal with interactions between infections (see below, 'Specific Aims'), and second, it suggests a mechanism for eventually merging our work with the product of existing efforts to organize and automate the recording and handling of medical record data. This latter development may in time make it possible for MYCIN to obtain a large part of the information it requires directly from such automated records, sharply reducing the number of questions it has to ask, and speeding up the consultation considerably.

Finally, several new capabilities make the system convenient to use, in anticipation of its evaluation in the clinical setting. Among these are the option of the user to type a comment about system performance at any time during the consultation. His comment is recorded in a special file which is reviewed periodically by our medical staff, and provides an ongoing opportunity for users to offer feedback aimed at improving the usefulness of the system. The user can also indicate his belief that the system has 'broken down' in some way and he is invited to describe the problem. His description is saved along with information about the current state of the program, so that our systems programmers can deal with the problem later.

8.5   Knowledge Acquisition

A preliminary knowledge acquisition program was completed in the middle of 1974, and demonstrated the feasibility of having

a physician teach the system new rules using a rather stylized subset of English. Building on the experience gained here, work began on a revised program designed to allow the user to examine and modify the program's knowledge and behavior as a single, unified action. This program was designed to make the explanation and knowledge acquisition capabilities available together, to make use of the fact that the nature of the explanations requested can give a clear hint about the content of a new rule. The program was also designed to advise the user about the effect of his rule on the original deficiency, indicating, for instance, whether or not it corrects the problem he noticed.

Work on a preliminary version of this new program was completed in 1976, making available a broad range of useful features enabling our clinical experts to add rules to the system without requiring that they have a knowledge of programming. If the expert finds that MYCIN's handling of a particular problem is at variance with his own expert knowledge, he can use the explanation capabilities to discuss the line of reasoning in use at that time, can add or modify rules in the knowledge base, and can determine the effects of the changes on MYCIN's subsequent performance. (Quality control is maintained on the overall system by regular meetings of our clinical and pharmacological experts who determine the 'official' MYCIN knowledge base.)

8.6   Technical Issues

As MYCIN's clinical capabilities have expanded, efficiency has improved as a result of a number of modifications to the system's technical capabilities. Early in our work, for instance, a comprehensive review and modification of the control structure was undertaken to improve efficiency and generality. The resulting program was both more direct, and faster.

More recently, modifications have been made so that the the large English dictionary can be kept on the disk and accessed only as needed, rather than keeping it in core, which slows down the system's response speed. The self documenting features of the program have also been improved to make them faster, and the system's interaction with the terminal has been made more uniform, to prepare for the time when different users of the system may have various different kinds of terminals.

8.7   Evaluation Activities

Since clinicians are likely to require documentation of MYCIN's competence and utility before seeking its advice, considerable time has been spent on evaluating the system and on

implementing a range of program features to support these efforts.

In the past two years we have obtained many useful suggestions from clinicians when the system was presented to several different conferences. In February 1975 it was presented to the Western Society for Clinical Research, in September 1975 to the International Symposium on Clinical Pharmacy and Clinical Pharmacology, and more recently (June 1976), it was presented to the Drug Information Association.

A large scale formal study and evaluation of MYCIN's performance was begun in January 1976. The same set of clinical data was provided to both MYCIN and a set of experts in infectious disease therapy. [Five of the experts were nationally recognized authorities in the field, the other five were clinical fellows in the Infectious Disease Division at Stanford. A complete list of names, titles and affiliations is found in the list of evaluators at the end of this report.] The judgments of the program and the experts were compared, and the experts were asked to evaluate MYCIN's performance.

To do this, we first designed a form to allow us to separate the variables requiring analysis. The parameters evaluated include

A. the 'quality' of the interaction - were any questions irrelevant or missing

B. the program's ability to determine organism identity

C. the program's ability to determine organism significance

D. the program's ability to select proper therapy

E. overall performance evaluation

F. potential impact as a clinical tool or teaching facility

The evaluation form was designed to be informative yet simple to complete. It was tested in a pre-evaluation trial run, then used for the formal study.

Consecutive patients with positive blood samples were evaluated for inclusion in the study by project personnel, until we obtained at least 10 patients for which MYCIN recommended therapy, and 15 patients overall (patients were rejected if they were outpatients when the sample was drawn, if they had a previous blood culture in the preceding seven days, or if they had a diagnosis of meningitis or infectious endocarditis.) For each of the patients accepted, a one to two page clinical summary

was prepared and combined with a summary of the laboratory test data as of the time when the first blood culture was obtained. This information was then used to obtain a therapeutic evaluation from MYCIN.

Each of the participating experts received a set of fifteen evaluation forms (one for each patient). Each form contained: (a) the clinical summary and lab data; (b) space for the expert to record his conclusions about the nature of the infection, likely causative organisms, and appropriate therapy; and (c) a transcript of the MYCIN consultation along with space for the expert to record his opinion of various aspects of MYCIN's performance. By presenting the information in this order, we obtained a therapeutic regimen from the expert based on the same information supplied to MYCIN. This allowed us to compare the expert's answers to MYCIN's, and also gave us the expert's opinion of the system's performance. In the past few months a sufficient number of the forms have been returned that we were able to do a preliminary analysis. The figures below are based on the nine (out of ten) which have been returned.

Since it is difficult to select a single number which summarizes performance, we have in general measured each of the parameters listed above in three ways: (i) the percent of instances in which the program was judged exactly correct, (ii) the percent of instances in which the program's performance was judged exactly correct or an acceptable alternative, and (ii) the percent of cases in which a majority of the experts judged its performance exactly correct or an acceptable alternative. By using all three measures, we obtain a range of figures which give a good picture of the program's performance.

All of these attempts to evaluate performance are complicated by the fact that (as expected) the experts' own choices about each patient were not unanimous. Thus, we cannot ask whether MYCIN's answers were 'correct' in any absolute sense, since there was no agreement on what constitutes 'correct'. Instead, we ask how often each individual expert rated the program's responses as correct. But given the variation among experts themselves, the program can never be expected to reach 100%, and depending on the extent of the intra-group variation, the absolute limit may in fact be much lower. Thus the ideal question to ask is Do experts rate MYCIN's performance correct at least as often as they rate each other's performance correct? This would give a good indication of how close the system's performance was to that of the group of experts as a whole.

We have been able to do this in a few isolated cases, but in general it requires more information than we were able to collect. This is discussed in more detail below, but in general terms the problem is that we were able to ask each expert for his choices for each patient, and ask him to rate MYCIN's choices. But, without a second round of questionnaires, which would ask

each expert to rate the acceptability of the other 9 experts'
responses,   we lack   direct information   about intra-expert
variability.   The figures below should be reviewed with this
caveat in mind.

A. 'Quality' of the interaction

To measure the first item, the experts were instructed to
mark any questions in the consultation which they felt were
irrelevant, and to note any questions which they felt were
omitted by the system. Overall MYCIN did quite well, as there
were no consultations in which a majority of the experts felt
that any particular question was irrelevant or omitted.   On the
average,   there were 0.53 questions judged irrelevant and 0.55
indicated as omitted.

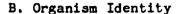Table I summarizes the next four measurements.

	% of instances MYCIN's first choice was identical to an expert's first choice	% of instances MYCIN's first choice was identical to, or was judged an acceptable alternative to an expert's first choice	% of cases MYCIN's first choice was identical to, or was judged an acceptable alternative by a majority of experts
ORGANISM IDENTITY	56.3% N = 414	75.6% N = 414	81.8% N = 11
ORGANISM SIGNIFICANCE	91.7% N = 36	NA	100% N = 4
THERAPY SELECTION	12% N = 99	75% N = 99	91% N = 11
OVERALL PERFORMANCE	17.0% N = 135	59.3% N = 135	60.0% N = 15

Table I. Summary of nine experts' responses to MYCIN's
          performance on 15 cases

B. Organism Identity

For organism identity, the experts were asked to rate each of MYCIN's selections as exactly correct (they agreed that the organism was likely to be present), an acceptable alternative (they had not chosen that organism, but agreed it might be present), or an unacceptable choice (they disagreed with its selection). Since 11 of the cases were not contaminants, and there was a total of 46 organisms chosen by the system, with 9 experts rating each of those choices we have an N of 414 for the first two columns and 11 for the third.

In 56% of the instances the system's choices were identical to the experts', 75% of them were either identical or acceptable alternatives, and in 82% of the cases, its results were acceptable to a majority of the experts.

In addition, the experts were asked to indicate which organisms they felt MYCIN had overlooked in its diagnosis. For the 11 non-contaminant cases, the experts indicated an average of only 0.35 organism identities that were overlooked by the system. In no case did a majority of experts feel that any particular organism had been overlooked, suggesting that even the 0.35 figure is a result of intra-expert variation.

C. Organism Significance

The first question on the evaluation form gave the expert a chance to indicate that he felt the patient did not need to be treated. The first column of the second row indicates the number of times the expert indicated no treatment was necessary for a case in which MYCIN also judged the organism to be a contaminant. (There is no number in the second column since we did not ask about a 'close call' on whether or not to treat. In addition, the measurement is based only on the contaminant cases, since in many of the cases where both MYCIN and the expert determined that treatment was necessary, they based that decision on different organisms. We felt that it would be misrepresentative to call these situations 'agreements'.)

As the figures show, in only three out of 36 instances was there any disagreement with the system's decision on whether or not to treat.

D. Therapy Selection

The expert was asked to select therapy for the organisms which he felt were likely to be present before looking at MYCIN's therapy recommendation. He was then asked to judge MYCIN's choice of therapy for that patient. Since MYCIN was selecting therapy for the organisms which it felt were present (which may have differed from those chosen by the expert), this provides a fundamental comparison of performance - it compares therapy

selection performance of the two when they are faced with the same clinical situation.

This comparison is a difficult one to make, since it is complicated by the difficulty noted above, of variability in the experts' performance and the need to judge MYCIN with respect to that variability. Looking only at exact agreements (i.e., two identical therapies) produces the figure in the first column, which indicates that 12% of the time MYCIN's recommendation was identical to that of an expert. Comparing each expert's therapy choice with the other 8 indicates that 35% of the time (N= 396) any pair of experts chose identical regimens. The experts were also asked to judge whether MYCIN's therapy was an acceptable alternative (if it was not identical to their own), producing the figure in the second column. This indicates that it was either identical, or they felt it was an acceptable alternative 75% of the time. (Unfortunately, we have no reliable way of judging the intra-expert variability here, without a second round of questionnaires which asked each expert to rate the acceptability of the other experts' choices.) [As an alternative, we have attempted to develop a measure of how 'far apart' two non-identical regimens are. But the problem is difficult: for example, for gram negative rods with salmonella most likely, is gentamycin and chloramphenicol 'very different' from gentamycin and ampicillin? We have been working on a 'drug metric' to solve this problem, attempting to base the difference between two drugs on factors like organism susceptibility, toxicity, and drug efficacy, but this work is still in progress.]

The figure in the third column gives a crude overall measure of therapy selection performance, and indicates that in 91% (10 out of 11 cases), a majority of the experts rated MYCIN's regimen as either identical to their own or an acceptable alternative.

[The evaluation form also asked each expert to choose a regimen for the organisms which MYCIN had selected. The intent here was to compare the system's performance against the expert when both were faced with the same set of organisms (rather than compared with the same clinical situation, as above). Unfortunately, inconsistent answers on the part of the experts indicated that they were not answering the question according to the instructions. It appeared that they were not able to suspend their own judgments about organism identity sufficiently to select a regimen based on MYCIN's organisms alone. For this reason, we believe the data to be unreliable, and have not included it here.]

E. Overall Performance

At the end of each evaluation form, the expert was asked to rate the system's overall performance as either excellent, good, fair, or poor. The first two columns of the last row indicate

that 17% of these evaluations were 'excellent', and almost 60% were either 'excellent' or 'good' (only 13% were 'poor'). In 60% of the cases (9 out of 15), a majority of the experts felt that MYCIN's overall performance was either 'excellent' or 'good'.

F. Present Utility and Future Potential

Finally, after completing the entire set of 15 patients, each expert was asked to rate MYCIN's present utility and future potential as a clinical tool and as an educational tool, rating it as having 'considerable', 'some', or 'no' potential. The table below summarizes their response.

Evaluation of Present Utility

	'considerable'	'some'	'none'
clinical tool	11%	67%	22%
educational tool	11%	89%	0%

Evaluation of Future Potential

	'considerable'	'some'	'none'
clinical tool	11%	89%	0%
educational tool	67%	33%	0%

Table II. Opinions of 9 experts on MYCIN's present utility and future potential

To aid these evaluation efforts, we have also implemented a number of useful features in the system. For instance, MYCIN now keeps continuing statistics of the use of rules in its knowledge base. This will help us to monitor its long term performance, to study the interrelationship between rules, and perhaps detect automatically any inconsistencies or gaps in the knowledge base.

We have also designed and implemented a mechanism for 'on-line' evaluation. At the end of each consultation, the system asks a few questions about the quality of its performance from the clinicians who are using it. This interchange will be brief to avoid being a burden to the user, but it is expected to represent an important addition to the other evaluation efforts.

It will, for instance, make possible a new form of evaluation of the system. Rather than using a series of 'prepackaged' cases as was done in our initial evaluation, the next stage will be carried out using information entered at a

terminal by the evaluator. The participating panel of experts will be selecting patients in areas covered by the MYCIN knowledge base, and will engage in a dialogue with the system about those patients. Following completion of the session, the on-line evaluation feature will ask questions about system performance, and the responses will be tabulated and evaluated on-line by appropriate biostatistical programs. Specific recommendations which may point out problem areas in the consultation will be reviewed by our staff. By this process we expect to be able to maintain a continuing evaluation of MYCIN's capabilities in various areas, and pinpoint specific areas where performance is suboptimal.

STAFFING

Infectious Disease
Dr. Stanton Axline, MD	6/74 to present	co-prin. invest.
Dr. Victor Yu, MD	9/75 to present	research affiliate
Dr. Frank Rhame, PhD	9/74 to 9/75	research affiliate
Dr. Edward Shortliffe, PhD,MD	6/74 to 6/76	research assistant

Clinical Pharmacology
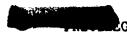|		
Dr. Stanley Cohen, MD	6/74 to present	prin. investigator
Dr. Robert Blum, MD	6/76 to present	research affiliate
Ms. Sharon Wraith, BS Pharm	6/75 to present	research associate
Dr. M. Goldberg, MD	9/75 to 9/76	research affiliate
Dr. Rudolfo Chavez-Pardo, MD	9/74 to 9/75	research affiliate

Computer Science
Dr. Bruce Buchanan, PhD	6/74 to present	investigator
Dr. Randall Davis, PhD	6/74 to present	research associate
Ms. A. Carlisle Scott, MS	6/74 to present	sci. programmer
Mr. William van Melle, MS	6/74 to present	research assistant
Dr. Cordell Green, PhD	6/74 to 6/75	asst. professor

Panel of Experts Participating in the 1976 Evaluation

National Experts

Dr. Dennis Maki, Chief of Infectious Disease, University of Wisconsin Hospital

Dr. John McGowan, Assistant Professor of Medicine, Infectious Disease Division, Grady Memorial Hospital, Atlanta, Ga.

Dr. Allan Kaiser, Chief of Infectious Disease, Vanderbilt Hospital

Dr. William Schaffner, II, Associate Professor of Medicine, Vanderbilt Hospital

Dr. Harvey Elder, Chief of Infectious Disease, Associate Professor of Medicine, Loma Linda University

Local Experts (and their current positions)

Dr. John Galgiani, Postdoctoral Fellow in Infectious Disease, Stanford Medical Center

Dr. Larry Lutwick, Postdoctoral Fellow in Infectious Disease, Stanford Medical Center

Dr. Rudy Johnson, Assistant Professor of Medicine, Vanderbilt University

Dr. Jerome Hruska, Assistant Professor of Infectious Disease, University of Rochester

Dr. Stanley Deresinski, Assistant Professor of Infectious Disease, University of South Florida

9   Appendix B: Hardware Announcement

Say DEC Readies 1st Unit In 32-Bit Computer Line

By RON ROSENBERG

MAYNARD, Mass. — Digital Equipment is reportedly readying the first of an expected new family of 32-bit computers that will be software-compatible to its 16-bit high-end PDP-11/70 and will utilize many of the performance features of its more expensive DECsystem 20 large computer.

Code named "VAX," the new computer could be introduced as early as October, according to sources, and be priced in the PDP-11/70 range but below the large DECsystem 20 starting price of $250,000. The system would initially compete against Interdata's 8/32 and System Engineering Laboratories' 32/75 machines. Both firms are currently the major suppliers of 32-bit systems.

Sources claim the key to the new DEC machine is its ability to run PDP-11 software using an emulation mode slightly slower than the PDP-11/70. They also note that VAX will utilize many of the DECsystem 20 features, such as a mass bus with five unibus ports. Machine throughput is reportedly between 10 and 25

(sidebar, vertical:) ELECTRONIC NEWS, MONDAY, APRIL 25, 1977   2
(sidebar, vertical:) ELECTRONIC NEWS, MONDAY, APRIL 25, 1977

Continued from Page One

megabytes per second.

The system is said to use emitter coupled logic (ECL) to achieve speeds approaching the DECsystem 10, DEC's largest and most expensive system.

DEC reportedly has launched. maintenance, manufacturing and test training at the company's leased Salem, N.H., facilities, not far from a major manufacturing center DEC is constructing. The new machines, sources claim, will not have DEC in-house-developed 32-bit software at the VAX introduction this fall. However, the main features of the instruction set are expected to be similar to IBM's 360 approach.

While Digital Equipment declined to comment on the new 32-bit system, it has been learned that DEC has made several presentations of the new system to Bell Laboratories' Holmdel, N.J., switching center and, reportedly, to several other large customers.

The new machine, according to several Wall Street sources, could be DEC's next generation of small computers designed to compete against the expected inroads of IBM's Series/1 and, to some extent, its mainframes. They cite how the small computer industry is approaching capacity performance with 16-bit architecture.

Industry sources said that 32-bit machines offer more direct memory addressing, doubling of the instruction length and a dramatic increase in peripherals and other input/output devices on a large 32-bit system. Interdata's 8/32 can directly address one megabyte of memory, considerably larger than the PDP-11/70.

The smallest DECsystem 20 is priced at $250,000 and the biggest is $400,000. The basic "VAX" system is expected to be considerably less than the former.

Industry sources noted that DECsystem 20, introduced less than 16 months ago, was designed to "bridge the gap" between the 16-PDP-11 and the DECsystem 10 (EN, Jan. 19, 1976). It is aimed to expand DEC's mainframe

business with a machine that has a CPU with 370/145 performance, but the price range of a 370/115-125. The 20/40 model also effectively replaces the earlier DECsystem 1040.

The move to 32-bit architecture has been rooted in the PDP-11/70 which has data transfer paths that would also be employed in a new machine, one source noted, adding that the PDP-11/70 is designed with a mass bus architecture.

10   REFERENCES

10.1   MYCIN PUBLICATIONS

Shortliffe, E H, Axline, S G, Buchanan, B G, Merigan, T C, Cohen, S N. An artificial intelligence program to advise physicians regarding antimicrobial therapy, Computers and Biomedical Research, 6:544-560 (1973).

Shortliffe, E H, Axline, S G, Buchanan, B G, Cohen, S N, Design considerations for a program to provide consultations in clinical therapeutics, Presented at San Diego Biomedical Symposium 1974 (February 6-8, 1974).

Shortliffe, E H, MYCIN: A rule-based computer program for advising physicians regarding Antimicrobial therapy selection, Thesis: Ph.D. in Medical Information Sciences, Stanford University, Stanford CA, 409 pages, October 1974. Also, Computer-Based Medical Consultations: MYCIN, American Elsevier, New York, 1976.

Shortliffe E H MYCIN: A rule-based computer program for advising physicians regarding antimicrobial therapy selection (abstract only Proceedings of the ACM National Congress (SIGBIO Session), p. 739, November 1974. Reproduced in Computing Reviews 16:331 (1975).

Shortliffe E H, Rhame F S, Axline S G, Cohen S N, Buchanan B G, Davis R, Scott A C, Chavez-Pardo R, and van Melle W J MYCIN: A computer program providing antimicrobial therapy recommendations (abstract only). Presented at the 28th Annual Meeting, Western Society For Clinical Research, Carmel, CA, 6 Feb 1975. Clin. Res. 23:107a (1975). Reproduced in Clinical Medicine, p. 34, August 1975.

Shortliffe, E H and Buchanan, B G, A Model of Inexact Reasoning in Medicine, Mathematical Biosciences 23:351-379, 1975.

Shortliffe, E H, Davis, R, Axline, S G, Buchanan, B G, Green, C C, Cohen, S N, Computer-based consultations in clinical therapeutics: explanation and rule acquisition capabilities of the MYCIN system, Computers and Biomedical Research, 8:303-320 (August 1975).

Shortliffe E H, Axline S, Buchanan B G, Davis R, Cohen S, A computer-based approach to the promotion of rational clinical use of antimicrobials, International Symposium on Clinical Pharmacy and Clinical Pharmacology, Sept 1975, Boston, Mass. (invited paper)

Shortliffe E H, Judgmental knowledge as a basis for computer-

assisted clinical decision making, Proceedings of the
1975 International Conference on Cybernetics and Society,
pp 256-7, September 1975.

Davis R, King J J, An Overview of Production Systems, Machine
Intelligence 8: Machine Representations of Knowledge (eds
E W Elcock and D Michie), John Wylie, April 1977. (Also
Memo HPP-75-7, Stanford University, October 1975).

Davis R, Buchanan B G, Shortliffe E H, Production rules as a
representation for a knowledge-based consultation system,
Artificial Intelligence, Vol 8, No 1 (February 1977).
(Also Memo HPP-75-6, Stanford University, October 1975).

Shortliffe E H, Davis R, Some considerations for the
implementation of knowledge-based expert systems, SIGART
Newsletter, 55:9-12, December 1975.

Scott A C, Clancey W, Davis R, Shortliffe E H, Explanation
capabilities of knowledge based production systems,
American Journal of Computational Linguistics, Microfiche
62, 1977. (Also Memo HPP-77-1, Stanford University,
February 1977).

Wraith S, Aikins J, Buchanan B G, Clancy W, Davis R, Fagan L,
Scott A C, van Melle W, Yu V, Axline S, Cohen S,
Computerized consultation system for the selection of
antimicrobial therapy. American Journal of Hospital
Pharmacy 33:1304-1308 (December 1976).

B.G. Buchanan, R. Davis, V. Yu and S. Cohen, 'Rule Based Medical
Decision Making by Computer,' (To appear in Proceedings
of MEDINFO.77, 1977).

Davis R, Applications of Meta Level Knowledge to the
Construction, Maintenance, and Use of Large Knowledge
Bases. Memo HPP-76-7, Stanford University, June 1976.

Davis R, Meta rules: content directed invocation, to appear, Proc
ACM Conf. on AI and Programming Languages, August 1977.

Davis R, Knowledge acquisition in rule-based systems: knowledge
about representations as a basis for system construction
and maintenance, to appear, Proc. Conf. on Pattern-
directed Inference Systems, May 1977.

Davis R, Interactive transfer of expertise: acquisition of new
inference rules, to appear, Proc. Fifth IJCAI, August
1977.

Davis R, A decision support system for medical diagnosis and
therapy selection, in "Data Base" (SIGBDP Newsletter),
8:58 (Winter 1977).

Davis R, Buchanan B G, Meta level knowledge: overview and
applications, to appear, Proc. Fifth IJCAI, August 1977.

10.2   OTHER REFERENCES

[1] Reiman H H, D'ambola J, The use and cost of antimicrobials in
hospitals, Arch Environ Health, 13:631-636 (1966).

[2] Kunin C M, et.al., Use of antibiotics: a brief exposition of
the problem and some tentative solutions, Anns Int Med,
79:555-560 (1973).

[3] Sheckler W E, Bennett J V, Antibiotic usgae in seven
community hospitals, J Amer Med Assoc, 213:264-267
(1970).

[4] Roberts A W, Visconti J A, The rational and irrational use of
systemic antimicrobial drugs, Amer J Hosp Pharm, 29:828-
834 (1972).

[5] Simmons H E, Stolley P D, This is medical progress? Trends
and consequences of antibiotic use in the United States,
J Amer Med Assoc, 227:1023-1026 (1974).

[6] Kagan B M, Fanin S L, Bardie F, Spotlight on antimicrobial
agents, JAMA, 226:306-310 (1973).

[7] Meyer A U, Weissman W K, Computer analysis of the clinical
neurological exam, Computers and Biomedical Research,
3:111-117, (1973).

[8] Warner H R, Toronto A F, Veasy L G, Experience with Bayes's
Theorem for computer diagnosis of congenital heart
disease, Anns NY Acad Sci, 115:558-567, (1964).

[9] Gorry G A, Barnett G O, Experience with a model of sequential
diagnosis, Computers and Biomedical Research, 1:490-507,
(1968).

[10] Edwards W, N = 1, diagnosis in unique cases, Computer
Diagnosis and Diagnostic Methods, (Jacquez, ed.), pp 139-
151, C C Thomas, Springfield, Illinois, (1972).

[11] Silverman H, A Digitalis Therapy Advisor, MAC-TR-143, E E
Department, Mass. Inst. Tech,(1974).

[12] Kulikowski C A, Weiss S, Safir A, Glaucoma diagnosis and
therapy by computer, Proc Annual Meeting of Assn for
Research in Vision and Opthalmology, (May 1973).

[13] Kunin C M, Tupasi T, Craig W A, Use of Antibiotics: a brief

exposition of the problem and some tentative solutions, Annals Int Med, 79:555-560, Oct 1973.

[14] Davis R, Applications of Meta Level Knowledge to the Construction, Maintenance, and Use of Large Knowledge Bases. Memo HPP-76-7 Stanford University, July 1976.

[15] Raiffa H, Decision analysis: introductory lectures on choices under uncertainty, Addison Wesley, 1968.